Journal of Green Engineering (JGE)

Volume-10, Issue-3, March 2020

An Efficient Botnet Detection Approach for Green IoT Devices Using Machine Learning Techniques

[1]M.Shobana and [2]S.Poonkuzhali

[1]*Research Scholar, Department of Information Technology, Rajalakshmi Engineering College, Chennai, India. E-mail:divyashobana.m@gmail.com*
[2]*Professor, Department of Information Technology, Rajalakshmi Engineering College, Chennai, India.*

Abstract

In this internet era, Internet of things (IoT) is expanding at an accelerating pace connecting billions of devices in our daily life. As more and more, green IoT devices are connected, securing IoT systems presents a number of unique challenges such as spoofing attacks, intrusions, denial of service (DoS) attacks, and distributed denial of service (DDoS) attacks, jamming, eavesdropping and malwares. This research work focus on malware detection. The majority of the malware which affects the IoT devices are botnets. These botnets are capable to produce a large amount of DDoS flood over its network. Recently Malware Detection using Machine Learning algorithms is gaining prominence to detect anomaly in the network traffic successfully. In this work, density based outlier detection technique is deployed to cluster botnet traffic and the normal traffic separately. Furthermore, these clusters are classified by SVM, Decision Tree and Naïve Bayes. The clustering and classification phase of this model is trained and tested using N-BaIoT dataset. The experimental result shows that almost 99.9% accuracy and can successfully detect MIRAI and BASHLITE attacks.

Keywords: Clustering, Classification, DBSCAN, IoT malware, outlier detection

Journal of Green Engineering, Vol. 10_3, 1053–1076.

Alpha Publishers

1 Introduction

The term INTERNET OF THINGS denotes the elevation of normal things into smart things. These smart objects include all sorts of domestic as well as industrial equipment. So these intelligent objects play a vital role in many major fields like military, vigilance, domestic and medical fields. The furious growth of this technology in the short span of time yields both merits and demerits to the corresponding field. Like how traditional computer was fully exposed to the internet before the implementation of suitable security framework, now these smart object tends to suffer the same issue, even worse condition is created for energy saving Green IoT devices [1]. This phenomenon will keep on raising because the count of IoT devices will reach up to 50 million in the year of 2020[2].The majority of the attacks launched to the IoT devices are DDoS attack. These kind of attacks are probably held by the malwares called Botnets. Botnets are the slaves which are hijacked and it is remotely controlled by groups or single bot controller. The main motive behind the bot controller is that creating flood of packets towards particular enterprise network to make it inactive for legitimate users of the site [3].

The main reasons for the green IoT devices to get targeted by the attacker for being their bots are due to the following reasons [4]:

- IoT devices does not have strong username/password so it is easy for the hacker to enter into it using brute force attack
- These devices have been online for 24X7, it is the main advantage for the hacker to launch mass DDoS attack without spending much cost
- Most of the green IoT devices are not directly controlled by humans to check its behavior at each moment

Recently many researchers have concentrated on the security breaches which are faced by the IoT network[5]. But only a few researchers have started to analyze about the malware that exist over the IoT network. There is always a war exist among the security researcher and hacker. Every time hacker keeps on changing their approach to establish the attack. Now IoT devices is the choice of them and still there are few challenges that has been exist to design a malware detection technique to the IoT devices viz[6].

- Green IoT devices are always said to be resource constrained, so the heavy weight algorithm can’t be used here.
- Green IoT devices work in a heterogeneous network with different protocols and algorithms. So the security algorithm should support all these different protocols.

So it is concluded that design of malware detection techniques to the IoT devices should be different from the existing traditional techniques. In this

work, the detection of botnets in the IoT environment is done by using outlier detection technique followed by machine learning. The reason behind to opt the outlier detection technique is to produce efficient results with less computation time [7]. Normally the existence of abnormalities or noise in the mass amount of data or dataset is considered as Outliers. The existence of outliers spoils the results produced by the particular corrupted dataset. These kind of outliers is detected using many simple statistical techniques and this technique is widely used in all areas. Recently it seeks attention in network security field. In this work, real time network traffic is analyzed by one of the outlier detection technique to spot out the outliers/anomaly. Since Green IoT devices are considered to be resource-constrained and low power consumption, the deployment of heavy weight classifiers with huge amount of training data will lead to slowing down the performance of IoT devices. In order to meet the above mentioned challenges, the clustering technique is adopted to cluster the incoming traffic of IoT devices without any prior knowledge about the previous network traffic. The pattern of the regular network traffic of the IoT devices is considered to be normal, then the occurrence of any deviation in the traffic is considered to be outlier (i.e.) malicious traffic. Then this pattern of outliers are trained and tested using classifiers. The most commonly used classifiers for malware detection has been deployed in this work.

The main contribution of this work follows as

- To perform dimensionality reduction of the IoT dataset using Principle component analysis
- To demonstrate DBSCAN(Density-based spatial clustering of applications with noise) algorithm to group the network traffic into many clusters with varied densities
- Resultant clusters are then classified into benign and attack traffic using three classifiers such as SVM, Naïve Bayes and Decision tree(ID3) based on its densities

2 State of Art

The recent work on malware detection or intrusion detection in IoT environment is always working with the help of machine learning and deep learning architecture. The behavioral attributes are considered as input parameters for these models and very specifically for botnet detection, statistical or dynamical attributes of network traffic are most preferable as input parameters. But these kinds of model suffer from accuracy due to non-availability of sufficient real time dataset for Green IoT devices. These kinds of problems can be overcome by doing the classification task followed by clustering technique. To the best of our knowledge, these kind of approach is not yet applied in IoT environment.

3 Related works

In Table 1, the work related to detect the malware in the IoT environment is discussed in elaborated manner. The approach used to detect botnets broadly classified into three type viz., signature based detection, honeypot detection and behaviour based detection. The existing work is designed using signature based detection, one of the traditional methods used to detect the malware. It provides high accuracy when compared to other technique, but it is prone to unknown attacks. Secondly honeypot detection like former method is also capable of capturing all attacks. But it fails to detect the attacks which are not accessible from it. Honeypot only detects the attack which comes directly to it. Moreover, if the honeypot is hacked by the attacker very soon it is easy for the hacker to cease the entire network attached to it. Thirdly behaviour based method, this model works according to the behaviour of the platform where it was implemented on it. In this survey, this model is expressed in three different ways such as simulation based method, machine learning and deep learning based method. In the case of simulation based method it works only on the emulated data rather than real time network data. Manually, it is impossible to generate all types of attack to test the designed model. Meidan et al [8] designed a model and trained only with benign traffic using deep autoencoders to capture malicious traffic. In this regard, the author has designed a separate model for each device. So it is easy to conclude that without prior network dataset, this model cannot run on that particular IoT device. Azmoodeh et al [9] proposed a model which operates on opcode sequence and trained using deep eigen space vector. Always deep learning model requires high amount of training data to produce high accuracy. Since IoT device is said to be resource constrained it is again an obstacle to implement these kind of model to implement in IoT device. Pajouh et al[10] presented a detection model for IoT device, it mainly deals with dimension reduction of the dataset. In this paper, the designed model is not trained and tested with the real time dataset. Nõmm et al[11] presented a common unsupervised model for all sorts of IoT device. The author has concentrated more on dimensionality reduction of the dataset. The approaches used for feature reduction are Hopkins statistics, Entropy and variance based feature reduction methods. Then the classifiers used in the classification process are SVM and Isolation forest. Shafi et al [12] proposed fog based SDN controlled intrusion detection and prevention systems for IoT devices. Author utilized multiple classifier system by the Recurrent Neural Network, Multi-Layer Perceptron and Alternate Decision Tree at the cloudlet or fog and it is further controlled by SDN manager. Nguyen et al [13] proposed a security architecture, namely *SeArch*, It works in a distributive manner and it is specifically for cloud based IoT gateways

which deploy intrusion detection techniques at three different nodes such as edge, fog and cloud. Hosseinpour et al [14] also presented an idea towards IDS which relies on three nodes namely edge, fog and Cloud. In this regard, author implements artificial immune system in a distributive way along those three nodes in order to discriminate normal traffic from malicious.

In this literature survey, it is clear that there is no previous research is based on an outlier detection algorithm which is applied for the IoT malwares. In this work, density based outlier detection technique is applied. DBSCAN (Density-based spatial clustering of applications with noise) is a density based clustering algorithm and it is capable of detecting collective outlier. Collective outlier is defined as the collection of points deviated from the entire dataset and this phenomenon is matched with the occurrence of malicious traffic in the IoT environment. Because the occurrence of malware attack cannot be considered as point or contextual outlier. In this algorithm, clusters are generated in any shape based on the values given in the dataset. It does not require the number of clusters to be formed from the user side.So this algorithm has the flexibility to work on a real time dataset. The clusters which lies in the lower density region is said to be outliers whereas clusters lies on the high density region is called as normal occurrence. The resultant clusters are then further classified into subcategory using three different machine learning classifiers such as SVM, Naïve Bayes and decision tree(ID3) these are common algorithms which are suitable for intrusion detection and it has been already proved in recent research[15] , So in this paper, these algorithms are adopted to implement in intrusion detection for IoT environment.

Table 1: To summarize the methodology used by the existing work

Author & year	Type of Approach	Technique Used	Deployed at	Input type	Captured Attack
Meidan et al [8] & 2018	Deep learning based approach	Deep Autoencoders	Network layer	Statistical Network attribute	Mirai and BASHLITE
Azmoodeh et al [9] & 2016	Deep learning based approach	Deep eigen space learning	IoT device	Opcode sequence	All possible IoT attacks
Abbas et al [16] & 2017	Signature based detection	-	IoT device	Signatures of system call of files	All known attacks
Habibi et al [17]& 2017	Simulation method	Whitelist based mitigation	router	Suspicious file	All possible IoT attacks
Pajouh et al[10] & 2016	Machine Learning based approach	Principal Component Analysis(PCA) Followed by KNN, Naive Bayes	Network layer	Network Traffic	U2R and R2L attack
Zhang et al[18] & 2015	Simulation method	-	Simulated IoT device	Network Traffic	IoT attack
Hughes et al [19] & 2014	Signature based detection	Logistic Regression	IoT Device	Signature of malware behaviour	All known IoT attack
Yin Minn pa pa [20] & 2015	Honeypot Based Detection	telnet based emulation	IoT network	Network Traffic of Malicious attack	telnet based attacks
Dowling et al[21] & 2017	Honeypot Based Detection	Zigbee protocol based emulation	Wireless sensor network	Network Traffic of Malicious attack	IoT attack
Slay et al[22] & 2018	Machine Learning based approach	Decision Tree, Association Rule Mining, Artificial Neural Network and Naïve Bayes	Network layer	Network traffic	IoT Botnets
Dao et al[23] & 2017	Filter based Simulation method	Self organising map(SOM)	IoT device	IoT Traffic	DDoS attack
Ozcelik et al [24] & 2017	SDN and Fog based approach	Threshold Random Walk with Credit Based Rate Limiting (TRW-CB) and Rate Limiting.	Fog and IoT device	IoT Traffic	Mirai

4 Design of the proposed model

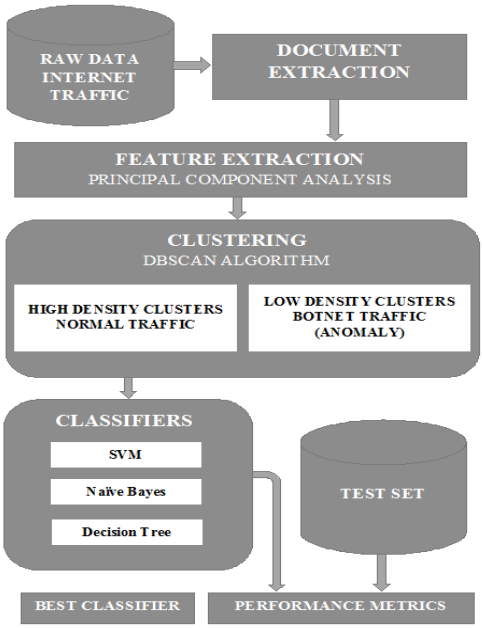


Figure 1. Framework of the Proposed Model

This model works on the network traffic generated by the nine different IoT devices. Meidan et al[8] generated this dataset which comprises of both normal and botnet attack traffic. Author induces two major botnet attack, namely Mirai and Bashlite. As illustrated in the Figure 1, this framework consists of three phases: feature extraction, clustering and classification.

4.1 Feature Extraction

This dataset consists of 115 network traffic attributes. Since this framework concerns only for IoT devices, the number of attribute must be reduced due to the space complexity. To do that in an efficient manner, Principal Component Analysis(PCA) is used for this step.

4.1.1 Principal Component Analysis

The goal of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a number of variables correlated with each other, while retaining the variation present in the dataset, up to the maximum extent. The same is done by transforming the variables to a new set of variables, which are known as the principal components and are orthogonal. Always the 1st principal component retains maximum variations that were present in the original components. The principal components are the eigenvectors of a covariance matrix, and hence they are orthogonal. In this regard, top two principal components (PC's) are considered for this work.

Steps involved in PCA algorithm

Input: N-Dimensional Dataset($D_N = d_1, d_2\ d_3 \ldots\ d_X$)
Output: Reduced Dataset($D_r = d_1, d_2\ d_3 \ldots\ d_y$)
Function_PCA
Begin
D_{N-1}`= remove the *class labels of* D_N
Compute mean values for each dimension (N-1), $D_{mean} = \frac{1}{X}\sum_{i=1}^{X} dX$.
Calculate the covariance matrix $D_{covariance} = \frac{1}{X}\sum_{i=1}^{X}(d - d`)(d - d`)$
Calculate the eigenvalues($e_1, e_2, e_3, \ldots\ e_x$) and eigenvectors($\lambda_1,\lambda_2,\lambda_3,\ldots\ldots\lambda_x$)
Arrange the eigenvectors in the given order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots\ldots\lambda_x$ and choose k eigenvectors with largest eigen values to form (**N-1)xk** matrix
Use this **(N-1)xk** matrix to transform the samples onto the new subspace D_r.
End

4.2 Clustering

Unlike all traditional machine learning techniques, clustering does not require any sample data for training purpose. Since storing the previous network traffic is again a backlog for resource constraint IoT devices, outlier detection technique based on clustering is deployed. Among all clustering algorithm, DBSCAN is capable of clustering the data in presence of noise. In this scenario, routers/gateways receive all sorts of IoT device's packets. Using DBSCAN algorithm, the traffic is clustered based on the density, whereas high density clusters are considered as normal traffic and low density clusters are considered as malware traffic.

4.2.1 DBSCAN

Density-based spatial clustering of applications with noise(DBSCAN) is an algorithm which is used to cluster the given data points in 2-dimensional space. These clusters are formed based on the distance measurement like Euclidean distance. DBSCAN requires two input parameters such as Minpts and Eps. Minpts denotes the minimum number of data points required to form the cluster. Eps denotes the eligible distance between the data points to consider them as neighbors.

The clusters obtained using DBSCAN clustering approach is analyzed to detect the botnet traffic. The clusters found in the low density region are identified as malicious traffic, whereas a cluster lies in the high density region implies as normal occurrence.

Steps involved in DBSCAN algorithm

Input: ε, minpts, *Reduced Dataset*($D_r = d_1, d_2\ d_3 \ldots\ d_y$)
Output: Set of clusters$C:(C_1, C_2\ C_3 \ldots\ C_m)$
Function_DBSCAN(ε, minpts, D_r)
Begin
E = Select a random data point.
Mark E as visited
N = Extract the neighbor points of E using ε
If |N| < minpts then
Mark E as outlier/anomaly
Else
C ← {E}
For each point E' ∈ N do
N ← N/E'
If E' is not visited then
Mark E' as visited
N' ← Extract the neighbor of C'
If |N'| ≥ minpts then
N ← N U N'
If E' is not member of any cluster then
C ← C U {E'}
Endif
Endif
End

4.3 Classification

The formed clusters are labelled as normal and malicious appropriately corresponding to its density of the clusters. The labelled clusters are classified using three machine learning classifiers viz., Support Vector Machine (SVM), Decision tree and Naive Bayes.

4.3.1 Support Vector Machine

SVM is a most commonly used learning algorithm for classification. This algorithm plots its n- feature in n-dimensional space. Then the algorithm will create a few hyper plane which cuts the data points into two classes. Among the few hyper planes, one of the best plane is selected on the basis of classification.

4.3.2 Decision Tree

Decision tree is a branch like structure algorithm. Here each attribute acts as a node and the relationship between the node acts as branches. Each branch is divided based on some decision or condition. This algorithm works well for the classification task. In this case, this algorithm is utilized here to classify malicious clusters from the normal clusters.

Steps involved in decision trees

Input : D_r ,Set of classified Instances
Output: Decision tree
*Function*_Decision Tree
Repeat
Maxgain ← 0
SplitA ← nill
e ← entropy(Attributes)
For all Attributes in D do
Gain ← InformationGain(a,e)
If gain > maxgain then
Maxgain ← gain
SplitA ← a
Endif
Endfor
Partition(D,SplitA)
Until all partitions processed
End

4.3.3 Naïve Bayes

Naïve Bayes algorithm works based on the Bayes theorem. It states that any two features classified is independent of each other. This classifier is well suited for high dimensional dataset. This algorithm performs the classification based on its previous occurrence and it is termed as prior probability.

5 Results & Discussion

The algorithm PCA and DBSCAN in proposed model was implemented on a Core i3 Laptop with 2.30 GHz CPU and 4 GB RAM using Rstudio version 3.5.1 software environment. Then the three classifiers were implemented using the WEKA software suite.

5.1 Dataset Description

The dataset N-BaIoT consists of traffic flows generated by the nine different IoT devices. All devices were subjected to two kinds of attacks, namely Mirai and Bashlite attack. Mirai attack has executed in five series stages they are Scan, Junk UDP: TCP, COMBO. Like Mirai, Bashlite attack also has five stages they are Scan, Ack, Syn, UDP and UDP plain.

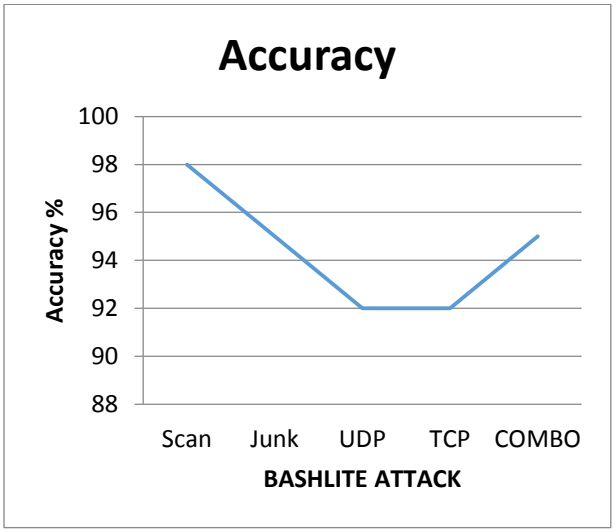
5.2 Outcome of Clustering

For the evaluation purpose, benign traffic of all devices were combined. In the same way, the dataset of Mirai and Bashlite attacks of all devices was combined, for example traffic of SCAN attack in Mirai is combined for all IoT devices. The clustering algorithm DBSCAN are performed on this combined dataset. In order to achieve high accuracy, the parameter Minpts and Eps has been tuned for many times. Table 2 lists the appropriate input parameter values for the corresponding botnet attack and it is very easy to conclude that Minpts values varies from 37 to 40 and Eps has lies between 0.12 and 0.20. The classification of low density and high density is based on its number of datapoints present in the clusters. Based on the experiments, it is observed that for various types of attacks, the number of datapoints in high density is from 110 to 5000. The clusters which have datapoints equal to or less than 50 is said to be low density clusters. On this basis the clusters are portioned into normal (low density clusters) and malicious (high density clusters). The accuracy of the cluster formation done by the DBSCAN algorithm is measured by the following equation (1) and the results are depicted in Figure 2 and 3:

$$\text{Overall accuracy rate} = \frac{\sum TP \ of \ all \ clusters}{number \ of \ instances} \quad (1)$$

Table 2: Performance and input parameters of DBSCAN

Type of Attack		Minpts	Eps	Number of clusters formed	Overall accuracy rate(%)
Bashlite	Scan	38	0.12	5	98
	Junk	40	0.15	8	95
	UDP	37	0.16	6	92
	TCP	40	0.20	6	92
	COMBO	40	0.15	7	95
Mirai	Scan	37	0.15	7	97
	Ack	40	0.10	6	95
	Syn	37	0.15	8	97
	UDP	37	0.15	7	98
	UDP PLAIN	37	0.16	8	97



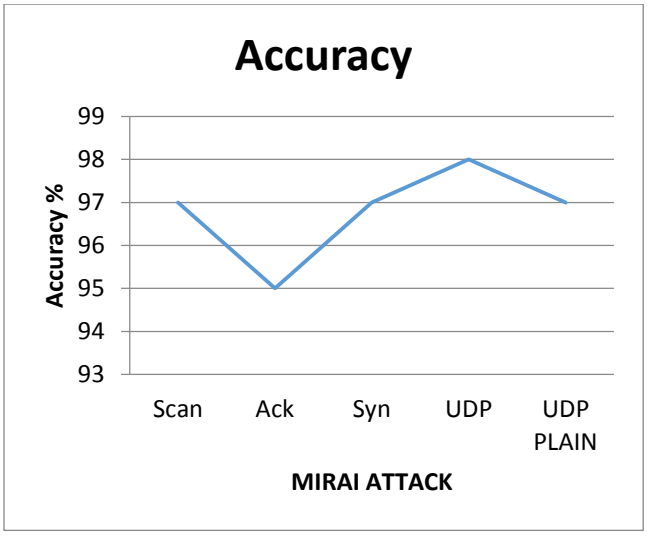


Figure. 2. Accuracy of DBSCAN algorithm to detect attack
(a) Bashlite , (b) Mirai

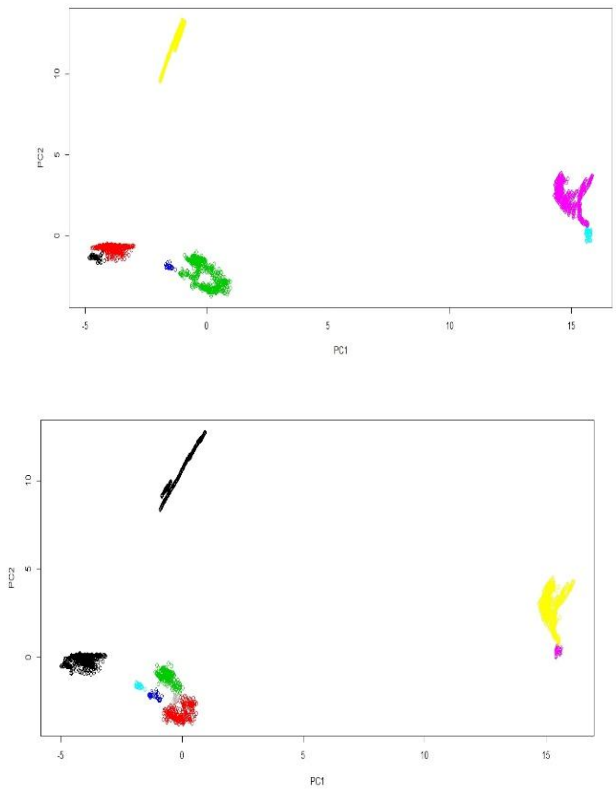


Figure.3. Clusters formed in SCAN attack using DBSCAN
(a) Bashlite (b) Mirai

In Figure 3 the clusters of SCAN attack in Mirai and Bashlite has been illustrated. In Figure 3(a) the cluster colored in blue and black is considered as anomaly whereas clusters colored as red, green and pink are considered as normal traffic. Likewise, in Figure 3 (a & b) the clusters colored in blue are considered as anomaly and the remaining clusters are considered to be normal. The accuracy of cluster formation using DBSCAN lies in between 92% to 98% whereas Emadi, H. S et al [25] used DBSCAN for clustering the sensor data and achieves accuracy of 95.1%. It is concluded that DBSCAN suits for clustering the real time network traffic of IoT environment.

5.3 Outcome of Classification

The performance of the three classifiers has been calculated by using the following metrics

- Accuracy

The accuracy of the classifier can be defined as the total number of correctly predicted instance in the given dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision

The Precision can be termed as the ability of the classifier should correctly label the malicious traffic as attack.

$$Precision = \frac{TP}{TP + FP}$$

- Recall

Recall can be defined as the number of correctly predicted normal traffic instances.

$$Recall = \frac{TP}{TP + FN}$$

- F-measure

F-measure can be defined as the weighted harmonic mean of precision and recall.

$$\mathrm{F-measure} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

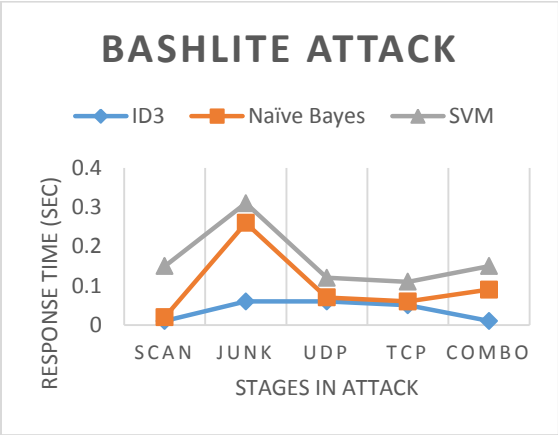
- Response time

The total time taken by the classifier to build the model accurately.

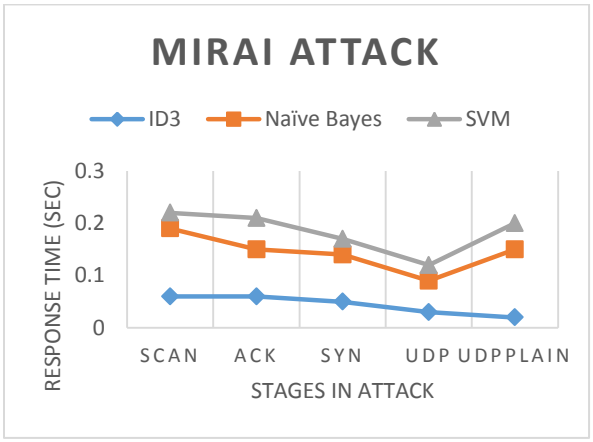
Figure 4 (a&b) illustrate response time taken by the three classifier for the two attacks. The response time for decision tree lies between the value from 0.01 to 0.06 and for Naïve Bayes the values varies from 0.01 to 0.2 whereas for SVM the time lies from 0.05 to 0.13. Figure 4a shows that for SCAN attack, ID3 and Naïve Bayes have lower value whereas SVM takes more time to detect. In the case of Junk and COMBO attack, ID3 has low response time than NB but SVM has higher response time than NB. Both classifier ID3 and NB detect UDP and TCP attack in minimum span of time than SVM. On the whole SVM is much sensitive to Bashlite attack than ID3 and NB.

Figure 4b illustrates that ID3 has less sensitive towards all the five stages of Mirai attack. In another point of view, all the three classifiers takes less response time for the 4th stage "UDP", this shows that Mirai attack can be detect at the 4th stage rather than its 1st stage whereas Bashlite attack can be in less time detected at the 1st stage itself.

In Figure 5(a) accuracy plot shows how each stage in Bashlite attack got classified accurately, SVM shows higher accuracy for SCAN than the other two classifiers. JUNK got similar accuracy around 99.50 for the three classifiers. For UDP & COMBO, the ID3 shows high accuracy about 99.81. For TCP, all the three classifier shows nearly similar value from 99.75 to 99.78. In figure 5b accuracy plot for Mirai attack has been demonstrated for each stage. In figure 6(a & b) the precision of both the attack is demonstrated. Precision values for all the stages is given for both the attacks are high ~0.995 except for UDP in both the attack are somewhat low around ~0.866. This scenario matches for all the three classifiers. Table 2 demonstrates the overall performance metrics of these algorithms which helps to find out the efficient classifier .By taking account into all the metrics, it is very easy to conclude that decision tree is very suitable to detect traffic of IoT Botnet.

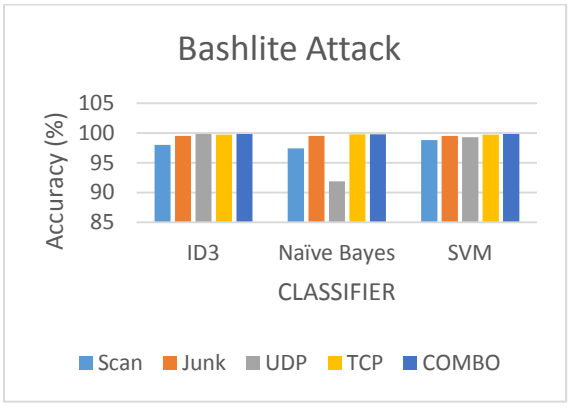


(a)

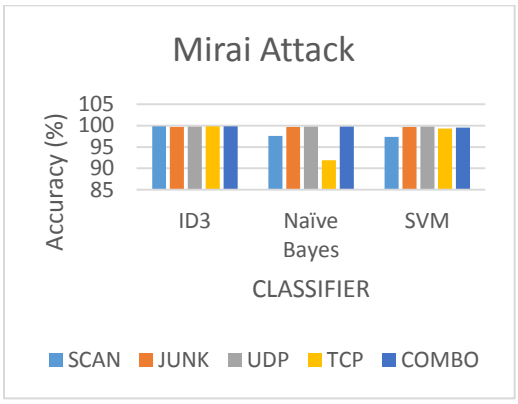


(b)

Figure. 4. Response Time to detect the attack
(a) Bashlite (b) Mirai

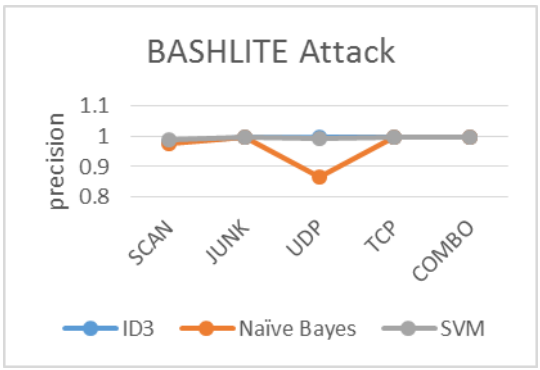


(a)



(b)

Figure. 5 Accuracy given by the three classifiers (a)Bashlite (b) Mirai



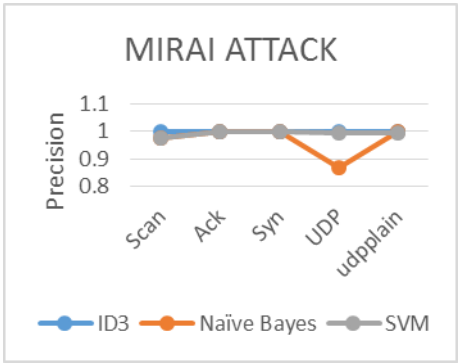


Figure 6 Precision for the three classifiers (a) Bashlite , (b) Mirai

Since decision tree produces high performance than other two classifiers for further discussions, its result will be considered accordingly. In comparison with other existing work such as Meidan et al [8] requires the detection time around 174 to 212 milliseconds but this works requires only around 60 milliseconds to detect all types of attacks. Table 3 compares the performance of the proposed model with existing work in terms of its accuracy. The accuracy produced by this model in the range between 98.05% to 99.81% to whereas the existing work achieves significantly lesser accuracy than this model. Table 4 demonstrates about the performance metrics for each classifier on detecting the attack.

Table 3: comparison of proposed model with existing work

Methodology		Methodology used	Accuracy (%)	Precision	Recall
Existing Work	hodo et al[26]	Multilayer perceptron	99.4	N/A	N/A
	pajouh et al [10]	Naïve Bayes and K-nn	84.82	N/A	N/A
	Alghuried, A[27]	IWC and Decision tree	97	0.78	0.98
	Bezerra et al[28]	OCSVM	96.65 to 99.33	0.95 to 0.96	0.99
	Nõmm, S et al[11]	OCSVM	86.18	0.50	N/A
		IF	95.65	0.90	
	Kumar, A et al[29]	Random Forest	88.8	0.86	1
		k-nn	94.44	0.92	1
		Gaussian Naïve Bayes	77.78	0.75	1
	Emadi, H. S et al [25]	SVM	97.1	N/A	N/A
	Shafi et al[12]	E3ML	~99	~0.98	~0.97
	Nguyen et al[13]	SeArch	95.50	0.95	0.95
	Hosseinpour et al[14]	Artificial Immune system	98.35	0.97	1
	Rathore et al [30]	ELM based Semi-supervised Fuzzy C-Means method(ESFCM)	86.53	86.59	86.11
Proposed work	Shobana et al	SVM	97.40 to 99.81	0.97to 0.99	0.97 to 0.99
		Naïve Bayes	91.88 to 99.80	0.86 to 0.99	0.91 to 0.99
		Decision Tree	98.03 to 99.81	0.98 to 0.99	0.98 to 0.99

Table 4 Performance metrics of classifiers on detecting the attacks

Classifier	Type of Attack		Recall	Precision	F-measure	ROC	Accuracy (%)
ID3	BASHLITE	Scan	0.980	0.980	0.980	0.984	98.03
		Junk	0.995	0.995	0.995	0.968	99.50
		UDP	0.998	0.998	0.998	0.988	99.81
		TCP	0.998	0.998	0.998	0.984	99.75
		COMBO	0.998	0.998	0.998	0.988	99.81
	MIRAI	Scan	0.998	0.998	0.998	0.988	99.81
		Ack	0.997	0.997	0.997	0.991	99.69
		Syn	0.998	0.998	0.998	0.984	99.75
		UDP	0.998	0.998	0.998	0.988	99.81
		udpplain	0.998	0.998	0.998	0.988	99.81
NAÏVE BAYES	BASHLITE	Scan	0.974	0.974	0.973	0.859	97.44
		Junk	0.995	0.995	0.995	0.965	99.50
		UDP	0.919	0.866	0.886	0.504	91.88
		TCP	0.998	0.998	0.998	0.986	99.78
		COMBO	0.998	0.998	0.998	0.988	99.80
	MIRAI	Scan	0.976	0.977	0.974	0.844	97.60
		Ack	0.997	0.997	0.997	0.978	99.69
		Syn	0.998	0.998	0.998	0.986	99.78
		UDP	0.919	0.866	0.866	0.504	91.88
		udpplain	0.998	0.998	0.998	0.968	99.78
SVM	BASHLITE	Scan	0.988	0.988	0.988	0.988	98.82
		Junk	0.995	0.995	0.995	0.976	99.50
		UDP	0.993	0.993	0.993	0.996	99.32
		TCP	0.998	0.998	0.998	0.992	99.75
		COMBO	0.998	0.998	0.998	0.994	99.81
	MIRAI	Scan	0.974	0.975	0.972	0.996	97.40
		Ack	0.997	0.997	0.997	0.991	99.69
		Syn	0.998	0.998	0.998	0.992	99.75
		UDP	0.993	0.993	0.993	0.996	99.32
		udpplain	0.996	0.996	0.996	0.998	99.56

6 Conclusion

In this paper, IoT botnet attack has been identified with less computation using density based outlier detection algorithm DBSCAN and its dimensionality reduction has been carried out using PCA. The formed clusters are accurately classified by the Decision tree when compared to other two classifiers. This model works on the basis of the present data rather than referring the previous network traffic and it improves the intelligence of the IoT devices to identify the malicious traffic from the normal traffic. Since

this model works on statistical network attribute, it is easy for this model to capture the unknown malware attack rather than the existing model.

References

[1] Hossain. M.M, Fotouhi. M, Hasan. R, "Towards an analysis of security issues, challenges, and open problems in the internet of things", IEEE World Congress on Services, New York, USA, pp. 21-28, 2015.

[2] Borgia.E, "The internet of things vision: Key features, applications and open issues", Computer Communications, Vol.54, No.1, pp.1–31, 2014.

[3] Angrishi. K, "Turning internet of things (iot) into internet of vulnerabilities (iov): Iot botnets", arXiv preprint arXiv:1702.03681, 2017.

[4] Bertino. E, Islam. N, "Botnets and internet of things security", Computer, Vol.1, No.2, pp.76-79, 2017.

[5] Xu. T, Wendt. J. B, Potkonjak. M, "Security of IoT systems: Design challenges and opportunities", International Conference on Computer-Aided Design ,USA , pp. 417-423, 2014.

[6] Zhang. Z. K, Cho, M. C. Y, Wang, C. W, Hsu, C. W, Chen, C. K, Shieh, S, "IoT security: ongoing challenges and research opportunities", International conference on service-oriented computing and applications , Matsue, Japan ,pp. 230-234, 2014.

[7] Alam. S, Dobbie. G, Koh. Y. S, Riddle. P, "Web bots detection using particle swarm optimization based clustering", IEEE congress on evolutionary computation (CEC) Beijing, China, pp. 2955-2962, 2014.

[8] Meidan. Y, Bohadana. M, Mathov. Y, Mirsky. Y, Shabtai. A, Breitenbacher. D, Elovici. Y, "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders",IEEE Pervasive Computing, Vol. 17, No.3, pp.12-22, 2018.

[9] Azmoodeh. A, Dehghantanha. A, Choo. K. K. R, " Robust malware detection for internet of (battlefield) things devices using deep eigenspace learning", IEEE Transactions on Sustainable Computing, Vol.4, no.1, pp. 88-95, 2018.

[10]Pajouh. H. H, Javidan. R, Khayami. R, Ali. D, Choo. K. K. R, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks", IEEE Transactions on Emerging Topics in Computing, ,Vol.7,no.2, pp.314-322, 2016.

[11]Nõmm. S, Bahşi. H, "Unsupervised Anomaly Based Botnet Detection in IoT Networks", International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, pp. 1048-1053, 2018.

[12]Shafi. Q, Basit. A, Qaisar. S, Koay. A, Welch. I, "Fog-Assisted SDN Controlled Framework for Enduring Anomaly Detection in an IoT Network", IEEE Access, Vol.6, pp.73713-73723, 2018.

[13]Nguyen. T. G, Phan. T. V, Nguyen. B. T, So-In. C, Baig. Z, Sanguanpong. S, "SeArch: A Collaborative and Intelligent NIDS Architecture for SDN-Based Cloud IoT Networks", IEEE access, Vol.7, pp.107678-107694, 2019.

[14]Hosseinpour. F, Vahdani Amoli. P, Plosila. J, Hämäläinen. T, Tenhunen. H, "An intrusion detection system for fog computing and IoT based logistic systems using a smart data approach", International Journal of Digital Content Technology and its Applications, Vol.10, No.5, pp.34-46, 2016.

[15]Thaseen. I, Sumaiya, and Chaswani Kumar. "Intrusion detection model using fusion of PCA and optimized SVM", International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Mysore, India, pp. 879-884, 2014.

[16]Abbas. M. F. B, Srikantha. T, "Low-complexity signature-based Malware detection for IoT devices", International Conference on Applications and Techniques in Information Security, pp. 181-189, 2017.

[17]Habibi. J, Midi. D, Mudgerikar. A, & Bertino. E, "Heimdall: Mitigating the internet of insecure things", IEEE Internet of Things Journal, Vol. 4, no.4, pp.968-978, 2017.

[18]Zhang. C, Green. R, "Communication security in internet of thing: preventive measure and avoid DDoS attack over IoT network". In Proceedings of the 18th Symposium on Communications & Networking, San Diego, CA, USA ,pp. 8-15, 2015.

[19]Hughes. K, Qu. Y, "Performance Measures of Behavior-Based Signatures: An Anti-malware Solution for Platforms with Limited Computing Resource", International Conference on Availability, Reliability and Security, Fribourg, Switzerland, pp. 303-309, 2014.

[20]Pa. Y. M. P, Suzuki. S, Yoshioka. K, Matsumoto. T, Kasam. T, Rossow. C, "IoTPOT: analysing the rise of IoT compromises", In 9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15),-2015.

[21]Dowling. S, Schukat. M, Melvin, H, "A ZigBee honeypot to assess IoT cyberattack behavior", Irish Signals and Systems Conference (ISSC), Killarney, Ireland, pp. 1-6, 2017.

[22]Slay. J, "Towards Developing Network Forensic Mechanism for Botnet Activities in the IoT Based on Machine Learning Techniques", International Conference of Mobile Networks and Management MONAMI, Melbourne, Australia, pp 30-44,2017.

[23]Dao, N. N, Phan, T. V, Kim.J, Bauschert. T, Cho. S, "Securing Heterogeneous IoT with Intelligent DDoS Attack Behavior Learning", arXiv preprint arXiv:1711.06041, 2017.
[24]Ozcelik. M, Chalabianloo. N, Gur. G, "Software-Defined Edge Defense Against IoT-Based DDoS", IEEE International Conference on Computer and Information Technology(CIT), Helsinki, Finland, pp. 308-313, 2017.
[25]Emadi. H. S, Mazinani. S. M, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks", Wireless Personal Communications, Vol.98, no.2, pp. 2025-2035, 2018.
[26]Hodo. E, Bellekens. X, Hamilton. A, Dubouilh. P. L, Iorkyase. E, Tachtatzis. C, Atkinson, R. "Threat analysis of IoT networks using artificial neural network intrusion detection system", International Symposium on Networks, Computers and Communications (ISNCC), pp. 1-6, 2016.
[27]Alghuried, A, "A Model for Anomalies Detection in Internet of Things (IoT) Using Inverse Weight Clustering and Decision Tree", Masters dissertation, Dublin Institute of Technology, 2017
[28]Bezerra. V. H, da Costa. V. G. T, Junior. S. B, Miani. R. S, Zarpelao, B. B, "One-class Classification to Detect Botnets in IoT devices", In SBSeg ,pp. 43-56, 2018.
[29]Kumar. A, Lim. T. J, "EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques", arXiv preprint arXiv:1906.09715, 2019.
[30]Rathore. S, Park, J. H, "Semi-supervised learning based distributed attack detection framework for IoT",Applied Soft Computing, Vol.72, pp.79-89, 2018.

Biographies

M.Shobana, Research Scholar pursuing Ph.D in Anna university. She completed her Master degree in SASTRA university on VLSI design. Her area of interest is Network security, malware detection and data mining. She published many research articles in international journals and conferences.

S.Poonkuzhali, Professor in the Department of Information Technology and Head of Centre for Assistive Devices and Technologies has been with Rajalakshmi Engineering College since August 2000. Her area of specialization is Web Mining, Outlier mining, Information Retrieval, Knowledge Management, Big Data Analytics and E-Learning. Authored 6 texts books and published more than 75 papers in various reputed conferences and in international journals. Received 6 Best Paper Awards for the oral paper presentation. Received International Paper Presenter Award from Computer Society of India. She is a recipient of Shri P.K. Award Memorial Best Faculty Award for Computer Science in Senior Category-2016 from Nehru Group of Institutions. Received Grants from CSIR, DST, CICS for presenting papers in International conference for oral presentation. Received Best Department award thrice under her leadership. Visited 8 countries for presenting papers, chairing sessions and as a plenary speaker. She is a reviewer for various refereed journals. She received grants to the worth of 90.5 Lakhs from various funding agencies for R & D activities. She had completed three funded projects from DST-TIDE, AICTE-RPS and CSI-MRP. Currently received recommendation for the project E-TLSID: Tool to impart Livelihood Skills for Intellectually Disable" from DST-TIDE. She is doing consultancy work for Sentinel Radiologist Solutions and Skill Council for Person with Disability. Life time Member in ISTE, CSI, IAENG and IACSIT. At present, Chairperson for Computer Society of India – Chennai Chapter. Currently, guiding 8 research scholars.